Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Artificial Intelligence 151 (2003) 213–225
Artificial Intelligence
www.elsevier.com/locate/artint

Research Note

Complexity of probabilistic reasoning in directed-path singly-connected Bayes networks

Solomon E. Shimony [a,*], Carmel Domshlak [b]

[a] *Department of Computer Science, Ben-Gurion University, Beer-Sheva 84105, Israel*
[b] *Department of Computer Science, Cornell University, Ithaca, NY 14853, USA*

Received 26 March 2002

Abstract

Directed-path (DP) singly-connected Bayesian networks are an interesting special case that, in particular, includes both polytrees and two-level networks. We analyze the computational complexity of these networks. The prediction problem is shown to be easy, as standard message passing can perform correct updating. However, diagnostic reasoning is hard even for DP singly-connected networks. In addition, finding the most-probable explanation (MPE) is hard, even without evidence. Finally, complexity of nearly DP singly-connected networks is analyzed.

Keywords: Probabilistic reasoning; Bayes networks; Complexity; Singly-connected DAGs

1. Introduction

Using Bayes networks (BNs) to model uncertain knowledge, and the problem of performing inference in this model, are of major interest in both theoretical and applied AI research [21]. As inference over BNs is hard in the general case [4,6,28], complexity analysis of sub-classes of Bayes networks is of extreme importance: knowledge of the exact frontier of tractability impacts heavily on the type of Bayes networks one may wish to acquire from experts or learn from data [9]. Additionally, numerous inference algorithms on BNs essentially use a reduction to one of the known tractable classes, in order to perform inference. The reduction is usually exponential in some aspect of the problem instance. For example, conditioning algorithms [10,13,21] use a cutset whose removal reduces inference

* Corresponding author.
E-mail addresses: shimony@cs.bgu.ac.il (S.E. Shimony), dcarmel@cs.cornell.edu (C. Domshlak).

0004-3702/$ – see front matter
doi:10.1016/S0004-3702(03)00110-3

into a number of (easy) inference problems on polytrees—where the number of polytree inference problems is exponential in the cutset size. Similarly, clustering schemes [14,17] aggregate nodes into macro-nodes so as to create a tree, where problem reformulation cost is exponential in the number of nodes in each macro-node.

Broadening the known classes of tractable Bayes networks is important, as it can improve inference performance for a given problem instance in two ways: (a) if a given problem instance happens to belong to the new class of tractable problems, a specialized algorithm for this class can be used, and (b) the cost of transforming a problem instance to the new class (e.g., a prediction problem on directed-path singly-connected networks, discussed in this paper) may be lower than in previously known reductions (e.g., to a polytree). In this paper, we analyze the complexity of directed-path (DP for short) singly-connected Bayes networks, for various inference problems, and present efficient algorithms for handling the tractable cases.

As excellent introductions to Bayes networks abound [3,19,21], we restrict this introduction to defining our notation, as well as to overviewing the standard inference problems on BNs. A Bayes network $\mathcal{B} = (G, P)$ represents a probability distribution as a directed acyclic graph G (see Fig. 1), where its set of nodes V stands for random variables (assumed discrete in this paper), and P, a set of tables of conditional probabilities (CPTs)—one table for each node $X \in V$. For each possible value x of X, the respective table lists the probability of the event $X = x$ given each possible value assignment to (all of) its parents. Thus, the table size is exponential in the in-degree of X. Usually it is assumed that this in-degree is small. Otherwise, representation of the distribution as a Bayes network would not be a good idea in the first place. The joint probability of a complete state (assignment of values to all variables) is given by the product of $|V|$ terms taken from the respective tables [21]. That is, with $\Pi(X)$ denoting the parents of X in G, we have:

$$P(V) = \prod_{X \in V} P\big(X \mid \Pi(X)\big).$$

Directed-path singly-connected Bayesian networks are defined as networks where for every pair of nodes (s, t) in the directed acyclic graph (DAG) of the network, there is at most one *directed* path from s to t. The notion defined here, is *not* the same as polytree topology, where the requirement is that there be at most one path from s to t in the *underlying undirected graph*, i.e., the underlying undirected graph is a tree. For example, the network in Fig. 1(a) is a polytree, while the network in Fig. 1(b) is *not*, even though

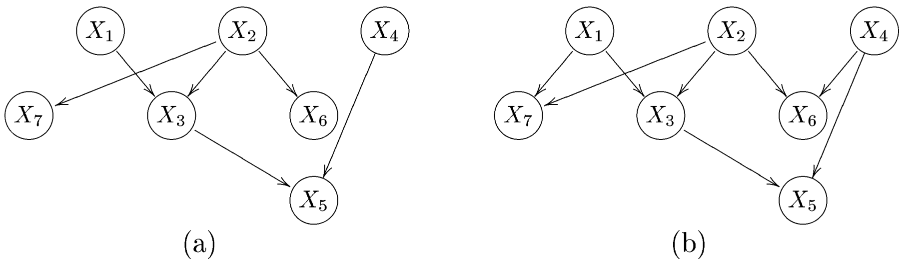

Fig. 1. Single-connectedness: (a) Polytree; (b) DP singly-connected graph.

it *is* DP singly-connected.[1] Reasoning on polytrees is known to be easy [16]—this is the most well known easy sub-class of Bayes networks. Clearly, from the above definitions, all polytrees are DP singly-connected, but not vice versa. Another important class of DP singly-connected DAGs, used in BNs in some applications [22], are all 2-level networks—which are singly-connected because all paths have length 1. Obviously, 2-level networks can have essentially unbounded tree-width [29], which would potentially place them in the "inference is hard" category, even for bounded in-degree.

Probabilistic reasoning (inference) is usually in one of two forms: *belief updating*, and *belief revision* [21]. In either case, a distinction can be made between a problem with *evidence*, which is a partial assignment $\mathcal{E}$ to some of the variables (presumably *observed* values for some of the variables), and a reasoning problem with no evidence. The belief updating problem is: compute marginal distributions for all the variables given the evidence, i.e., compute $P(X = x \mid \mathcal{E})$ for all $X \in V$ and for each possible value x of X. Belief revision has two variants. The first, also called most probable explanation (MPE), is finding the complete assignment A to all the variables that maximizes $P(A \mid \mathcal{E})$. The second, called MAP, is finding an assignment A to only *some* of the non-evidence variables that maximizes $P(A \mid \mathcal{E})$. All of the above problems are known to be NP-hard in the general case [4,6,28]. For the restricted case of polytree topology, belief updating and MPE are known to be easy [16], but MAP was recently shown to be hard [20].

Some versions of the null evidence problem are easy to approximate [6], but the problem is still hard in the general case. Similar results are shown here for DP singly-connected networks. We make the additional distinction between "diagnostic" problems, leading to a hard problem, and "predictive" problems, leading to an easy problem for belief updating. We observe that belief revision is in some sense harder than belief updating, as for DP singly-connected networks obtaining the null evidence MPE is NP-hard. The latter is somewhat surprising, given that some papers actually use sets of high-probability assignments to do approximate belief updating [23,24,26], with positive empirical results.

The rest of the paper is organized as follows. Section 2 shows that belief revision for DP singly-connected networks is hard, and that belief updating is hard for diagnostic problems. Section 3 shows how one can use a slight variant of the message passing algorithms [16] to compute marginal probabilities with no evidence in DP singly-connected networks. Expanding the scheme to more general prediction problems is discussed. Section 4 discusses belief updating for prediction problems in "almost" singly-connected networks.

2. DP singly-connected networks—hard problems

We show here that belief updating and belief revision (MPE) on DP singly-connected Bayes networks are hard in the general case.

[1] In the graph-theory literature (e.g., see p. 485 in [5], or [15]), what we call "directed-path singly-connected" is called simply "singly-connected", a more natural term. However, in the probabilistic reasoning community, numerous papers, e.g., [18,21,24,31] use the term "singly-connected networks" interchangeably with "polytrees"—we thus opted for the term "directed-path singly-connected" in order to avoid a possible confusion of these terms.

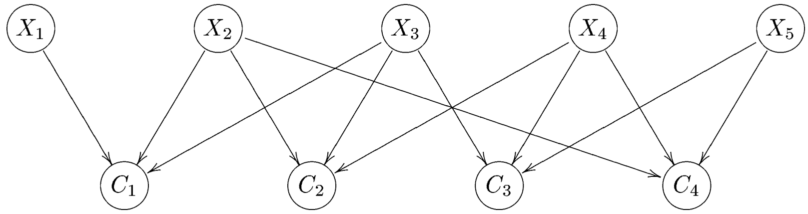

Fig. 2. DP singly-connected Bayes net for the proof of Theorem 1.

2.1. Diagnostic belief updating is hard

The fact that belief updating on DP singly-connected networks is hard is not surprising, since all 2-level networks are DP singly-connected, and it is "common knowledge" that these 2-level network problems are hard. Specifically, some diagnostic systems [22] use 2-level networks, and except for the well-known easier cases, such as monotonic abduction [25], performing diagnosis was understood to be hard. However, we have failed to find an actual proof of this result for DP singly-connected networks, or even for the commonly used 2-level networks. Below, we present a novel proof for 2-level networks.

Theorem 1. *Belief updating in* 2*-level networks* (*and thus also in DP singly-connected networks*) *is* NP*-hard.*

Proof. We use a reduction from unique SAT, which is the following problem.

Definition 1. Unique SAT (USAT) is: given a 3-CNF propositional formula that is known to have at most one model, is the formula satisfiable?

USAT is hard unless RP = NP, since it can be reduced from SAT by using a randomized polynomial-time algorithm [30]. Now, given a 3-CNF problem instance S, with clauses $C_1, \ldots, C_n$, each with literals $L_{i,1}, L_{i,2}, L_{i,3}$ over variables $X_1, \ldots, X_m$, create the following Bayes network (see Fig. 2). For each variable and each clause create a binary random variable—which we will denote by the same names, i.e., $X_1, \ldots, X_m$ and $C_1, \ldots, C_n$, respectively, for conciseness. The X_i variables have no parents, and have a prior probability of 0.5 of getting the value T (true). The C_i, called the *clause* nodes, each have at most 3 parents—exactly the variables mentioned in C_i. The conditional probability of $C_i = T$ given its parents is 1, except when the state of its parent variables makes $L_{i,1}, L_{i,2}, L_{i,3}$ all false, in which case conditional probability of $C_i = T$ is 0.[2]

Now, set as evidence—all C_i are true, and compute posterior marginals for all X_i. Let A be the assignment to all the variables, that assigns to each X_i the value x_i, where x_i is the value where the event $X_i = x_i$ has non-zero posterior probability. Clearly, if S is satisfiable, then, as it is uniquely satisfiable, assignment A is a model (and in that case, the

[2] We follow the reduction used in the top 2-levels in [4]. Unlike the earlier proof, we do not use the AND nodes below the clause nodes (Fig. 5), as the AND nodes violate DP single-connectedness.

events $X_i = x_i$ above will in fact have a posterior probability of 1). The converse holds, because all we need to do is to check whether A is a model, which takes linear time. If A is not a model, we know that S is unsatisfiable. Note that we actually need to perform the check, because if S is unsatisfiable, then the probability of the evidence is 0, and the posterior probability is undefined—A will be "garbage". □

The proof above uses a large number of evidence nodes. In fact, it seems that for DP singly-connected networks, belief updating complexity may well be at most exponential in the number of evidence nodes—but linear in the size of the network (assuming bounded in-degree). In addition, on examining the details of the proof, the following corollary is obtained, due to the fact that the reduction relies only on bounding the marginal probability away from 0:

Corollary 1. *Approximate belief updating in DP singly-connected networks within a constant finite ratio $\delta > 0$, is hard unless* RP $=$ NP.

In fact, we can also use a deterministic reduction from 3-SAT, which shows that belief updating is hard unless P $=$ NP, but it is not as clean as the above reduction. See Appendix A for the details.

2.2. MPE is hard

The fact that MPE with evidence is hard is well known, since for 2-level networks it can be reduced from the set-covering problem [22]. Proof that MPE without evidence is hard was shown in [28], but the reduction used a network that is not DP singly-connected.

However, by using the idea of additional "probability drain" nodes from earlier work [1,28], one can use the above reduction from SAT to show the desired result.

Theorem 2. *Finding MPE for DP singly-connected Bayes networks is* NP-*hard, even with no evidence. The respective decision problem "given K, is there a complete assignment with joint probability* (*not strictly*) *greater than K?" is* NP-*complete.*

Proof. By reduction from 3-SAT, as in Theorem 1, but here we do not need unique SAT. Given a 3-CNF formula S, construct the Bayes network as depicted in Fig. 2. Now, to each clause node C_i, add a binary variable D_i as a child—the "probability drain" nodes (Fig. 3). The conditional probability for these nodes is as follows: $P(D_i = T \mid C_i = T) = 1$, and $P(D_i = T \mid C_i = F) = P(D_i = F \mid C_i = F) = 0.5$.

Claim. *S is satisfiable just when there exists an assignment to all the Bayes network variables with a joint probability greater than or equal to $K = 1/2^m$, where m is the number of variables in S.*

The proof of the claim is constructive. Observe that any model M of S leads to one assignment A that has a joint probability of exactly $1/2^m$, as follows: simply set the variables X_i as in M, all the C_i to true, and all the D_i to true. Since the conditional

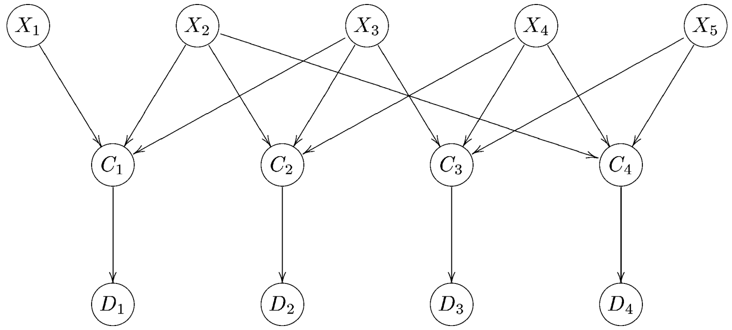

Fig. 3. DP singly-connected Bayes net for the proof of Theorem 2.

probabilities for all non-root variable assignments in A are 1, the joint probability is the product of the priors, which is $1/2^m$. Any other assignment B either violates the clause-literal constraints and thus has a joint probability of 0, or has some $C_i = F$. In the latter case, B must have a joint probability smaller by at least a factor of 2 (since any assignment, which we *must* make to D_i, introduces a factor of 0.5). Since the MPE decision problem is known to be in NP, the MPE decision problem without evidence on DP singly-connected networks is also NP-complete. $\square$

3. DP singly-connected BNs—prediction problems

In this section, we show that belief updating is easy for "prediction" problems. Denote the parents of node X by $\Pi(X)$, and by $D(\Pi(X))$ the set of all possible assignments on $\Pi(X)$. Likewise, denote by $\Pi^*(X)$ the set of ancestors (not necessarily immediate) of a node X, including X (the $*$ thus reads as "apply reflexive, transitive closure" to Π). The ancestor graph of a node X, denoted by $G^*(X)$, is the graph induced by X and its ancestors, i.e., consisting of $\Pi^*(X)$ and the edges from G connecting these nodes only. Likewise, we extend both of these notions to ancestors of a set of nodes. Note that, in DP singly-connected graphs, $G^*(X)$ forms an X-oriented polytree for every $X \in V$. That is, the underlying undirected graph of $G^*(X)$ forms a tree, and all the edges of $G^*(X)$ are oriented toward X.

Consider a Bayes network with no evidence. The marginal probability in any node, $X_i = x_{i,j}$, is expressed by:

$$P(X_i = x_{i,j}) = \sum_{A \in D(\Pi(X_i))} P(X_i = x_{i,j} \mid A) P(A). \tag{1}$$

When the Bayes net is DP singly-connected, then for each variable X, every two of its immediate predecessors Y and Z are d-separated [21] given no evidence. Thus, Eq. (1) can be evaluated as:

$$P(X_i = x_{i,j}) = \sum_{A \in D(\Pi(X_i))} P(X_i = x_{i,j} \mid A) \prod_{X_m \in \Pi(X_i)} P\bigl(X_m = A(X_m)\bigr). \tag{2}$$

If evaluated in topological order, this is equivalent to a limited form of standard null-evidence belief propagation for polytree Bayes networks (passing only π-messages [16]), and takes linear time. An alternate way to understand this property is by noting that when computing only the marginal probability of a node X in a DP singly-connected network, removal of the nodes that are barren with respect to X results in an X-oriented polytree.

Definition 2. Given a Bayes network $\mathcal{B} = (G, P)$, with evidence over a set of nodes E, a belief updating problem is called *strictly predictive* if the evidence nodes have no non-evidence parents in G.

Theorem 3. *Strictly predictive belief updating in DP singly-connected networks can be performed in time linear in the size of the network.*

The proof of Theorem 3 is straightforward—in DP singly-connected networks belief updating can done by fixing evidence node states, and then, as in the null evidence case, passing just the π messages.

Let us now consider a more general case, where the evidence nodes E in a DP singly-connected Bayes network $\mathcal{B} = (G, P)$ with nodes V, is an arbitrary set of nodes $E \subset V$. For every node $X \in V - E$, denote by $F(X)$ the ancestors of X on the "fringe" of the ancestors of the evidence nodes. Formally:

$$F(X) = \{U \mid [U \in \Pi^*(E)] \wedge [\exists W \in \Pi^*(X) - \Pi^*(E) \text{ s.t. } U \in \Pi(W)]\}.$$

Fig. 4 illustrates these notions: Fig. 4(a) is a graph of a DP singly-connected Bayes network. The shaded nodes in Fig. 4(b) stand for the evidence nodes, and $\Pi^*(E)$ are enclosed in the dashed box. For the illustrated query node X, the three nodes of $F(X)$ are emphasized by the dotted lines.

Definition 3. A belief updating problem is *weakly predictive* if the following (easy to check) conditions hold for the evidence node set E:

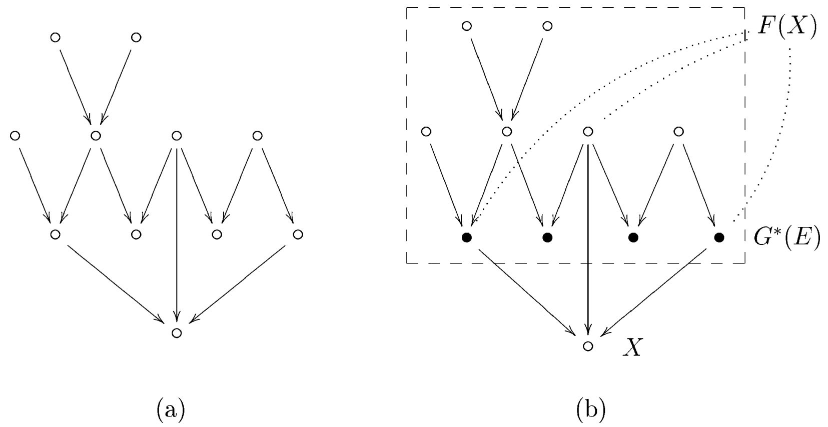

Fig. 4. DP singly-connected network and a weak prediction problem.

(1) $G^*(E)$ is a polytree, and
(2) For every node $X \in V$ not in $\Pi^*(E)$, all nodes in $F(X) - E$ are pair-wise d-separated in G given E.

For example, the evidence illustrated in Fig. 4(b) obeys the "weak prediction" condition, since (i) $G^*(E)$ is a polytree, (ii) X is the only node not in $\Pi^*(E)$, and (iii) there is only one node in $F(X) - E$.

Theorem 4. *Weakly predictive belief updating in DP singly-connected networks can be performed in time linear in the size of the network.*

Proof. First, (step 1) perform belief updating on $G^*(E)$ using the belief updating scheme for polytree Bayes networks [16]. The result is correct for nodes in $\Pi^*(E)$, since all other nodes can be considered barren nodes [27], i.e., nodes with no evidence or query node descendents. Now, (step 2) evaluate Eq. (2) in topological order for all nodes not in $\Pi^*(E)$, which takes linear time. The result is shown to be correct as follows. Let X be an arbitrary node in $G - G^*(E)$, and consider the subgraph $G'(X)$ induced by $\Pi^*(X) - \Pi^*(F(X)) + F(X)$. Since G is DP singly-connected, $G'(X)$ is an X-oriented polytree, with evidence only at the roots (nodes with no incoming edges). The root nodes of $G'(X)$ are independent given the evidence by construction (d-separation requirement on $F(X)$, and Definition 3), and their marginals (given the evidence) have already been computed in step 1. Thus, evaluating Eq. (2) results in correct probability updating for X. □

For example, see Fig. 4 illustrating a weak prediction problem. Most exact reasoning systems begin by removing barren nodes, and then, if the result is a polytree—use the linear-time algorithm. Otherwise, some (potentially exponential time) transformation to a tree-shaped network is made, followed by reasoning over the tree. In this case, there are no barren nodes, and although G is directed-path singly-connected, it is not a polytree. And yet, the evidence obeys the "weak prediction" condition, allowing probability updating for G in linear time using the above scheme.

We conclude this section by noting that it is not surprising that prediction tends to be easier than diagnosis, over the same type of topology. That is because evidence at root nodes blocks paths (in the sense of d-separation), and thus decreases the number of dependencies. However, evidence nodes that have ancestors tend to *unblock* paths, increasing the number of dependencies in the network.

4. "Almost" DP singly-connected networks

As the prediction problem is easy for directed-path singly-connected Bayes networks, it is of interest what happens when the network is nearly DP singly-connected. Two different definitions naturally come to mind. The first has to do with the cutset size, i.e., how many nodes need to be removed from the network in order to make it DP singly-connected? This is one standard notion used for inference in Bayes networks—and the runtime is exponential in the cutset size [21]. In our case, the scheme can be refined as follows: for each node X of interest, perform inference separately by:

(1) Creating $G^*(X)$, the ancestor graph of X.
(2) Find the cutset of $G^*(X)$ and perform belief updating using cutset conditioning [21].

The size of the cutset in the ancestor graph is usually much smaller than the one in the full graph, as has been noted in [7,8]. In fact, the above scheme is equivalent to taking advantage of the fact that the *predictive cutset* for X (using terminology from [7,8]) is smaller than the cutset for the full graph.

Another obvious possible extension is to bound the number of paths between pairs of nodes. A DAG is called *max-k-connected* if the number of different directed paths between any two nodes s and t in the graph is bounded by k. For example, max-1-connectedness is equivalent to directed-path single-connectedness. In some computational problems on graphical models in AI, for example, in planning, the complexity is "parameterized" in this manner by the directed connectivity of the graph [11]. Thus, we might expect the complexity of belief updating given predictive evidence to grow as a (possibly exponential) function of this connectivity bound, but polynomially in the size of the network. Surprisingly, this is not the case (Theorem 5):

Theorem 5. *Belief updating in max-k-connected networks is hard for all $k \geqslant 2$, even with no evidence.*

Proof. (For $k \geqslant 3$) The proof uses essentially the same reduction as in [4]—see Fig. 5, illustrating the graph of a Bayesian network reduced from a CNF satisfiability problem. In this reduction, again the top level has one binary-valued node per propositional variable, and the second level has one binary variable per clause. The rest of the network is set up such that node S is true with probability 1 just when all the clause nodes are true—otherwise, S is false with probability 1. (The nodes denoted A_i are AND nodes, used as mediators so that node S does not have a large number of parents.) It is known that the special case of 3-SAT, in which each variable participates in at most three clauses (denoted

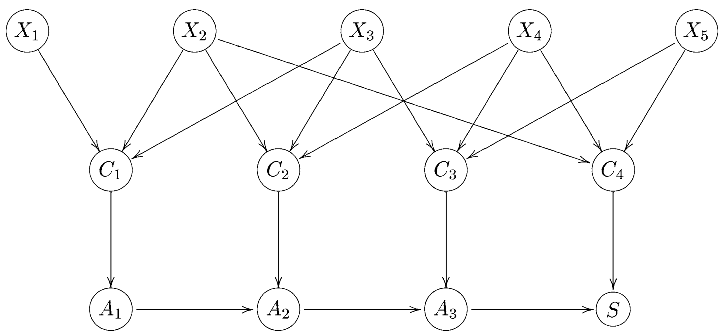

Fig. 5. Illustration of a BN constructed for 3-SAT.

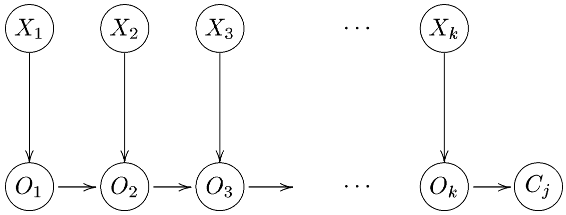

Fig. 6. Ladder of OR nodes between the X_i and the clause node C_j.

by 3μ-3CNF) remains NP-complete [12].[3] It easy to see that the Bayesian network created from an instance of CNF satisfiability, is max-k-connected if and only if each variable in the CNF formula participate in at most k clauses. □

In fact, in the above reduction, the probability of node S being true indicates the number of models of the CNF formula, since for m propositional variables, the number of models is $2^m P(S = T)$. The problem of counting models (denoted #SAT) is #P-hard, even in the restricted case of each variable appearing at most 3 times (#3μ-3CNF), and in fact was shown to be hard to approximate [25]. Hence, null-evidence belief updating for max-3-connected networks is also hard to approximate.

The proof for $k = 2$ is similar, but uses a reduction from #2μ-CNF (i.e., every variable appears at most twice, but the size of the clauses is not bounded by a constant), which is known to be #P-complete [2]. The reduction is the same as in Fig. 5, except for a minor technical problem, as follows. The fact that the size of a clause may be large does not allow us to add all the variables X_i as parents of the clause node, as the CPT will then have exponential size. The technicality is solved, again, by adding a ladder of nodes between the X_i and the clause node C_j, but this time the additional nodes are OR nodes (Fig. 6). Observe that there is a one-to-one mapping between edges in the graph in Fig. 5, and directed paths after adding the ladder in Fig. 6, and thus the graph representing a #2μ-CNF instance is max-2-connected, and the in-degree is bounded. Therefore, belief updating in max-2-connected networks is hard. However, since [2] has shown that #2μ-CNF is easy to approximate, it may well be the case that (null-evidence) belief updating for max-2-connected networks is also easy to approximate.

5. Discussion

The case of directed-path singly-connected Bayes networks is an important special case of Bayes networks, where predictive belief updating is easy, and belief revision, as well as diagnostic belief updating, is hard. The fact that in some cases a cutset-conditioning type algorithm may take advantage of the difference between "predictive cutsets" and the "standard" cutsets, has been observed before [7,8], and our results should shed further light

[3] Except for a very special case in which each clause contains *exactly* three literals.

on this issue. Designing an actual algorithm that uses this property, and evaluation of its performance on problem instances from applications remains for future work.

It is somewhat discouraging that in max-k-connected networks, null-evidence belief updating is hard for all $k \geqslant 2$—our initial intuitions were that the problem is exponential-time only in k. Nevertheless, the fact that approximate belief updating for max-2-connected networks is open, is also an interesting avenue for future research—both finding such an approximation algorithm, and possibly using it as a building block within a conditioning-type algorithm.

Acknowledgements

We would like to thank Amos Beimel for useful suggestions and pointers to some of the NP-hard problem variants used in the reductions. Research partially supported by an infrastructure grant from the Israeli Ministry of Science, and by the Paul Ivanier Center for Robotics and Production Management at Ben-Gurion University.

Appendix A

In Section 2, we showed a reduction from unique-SAT to belief updating in 2-level networks, where unique-SAT is hard unless NP = RP. Below is an alternate reduction, a deterministic reduction from 3-SAT. The topology of the reduction has the same structure as in Fig. 2, but in addition has a tree of binary-valued nodes S_i as in Fig. A.1, terminating in a "sentence" node S_m. The nodes X_i are *three-valued*, with the values $\{T, F, \perp\}$, where $\perp$ stands for "neither T nor F are consistent with the evidence".

Nodes X_i have uniform priors, i.e., $P(X_i = F) = P(X_i = T) = P(X_i = \perp) = \frac{1}{3}$. The nodes C_j are again binary nodes, but their CPT is now defined as follows. $C_j = T$ with probability 1 whenever its parents values are consistent with C_j, as before. If at least one of its parents has the value $\perp$, then $C_j = T$ with probability δ, where δ is a small positive constant defined below. Otherwise (i.e., none of the parents has a value of $\perp$, but the assignment does not satisfy the clause) then $C_j = F$ with probability 1.

The CPT of the S_i nodes is defined as follows: $P(S_i = T \mid S_{i-1} = T \text{ and } X_i \neq \perp) = 1$, and 0 otherwise. Prior for S_0 is $P(S_0 = T) = 1$. The evidence is: all C_j are T. Clearly, S_m can be true (with non-zero probability) only in instantiations where all the X_i have a non-$\perp$ value. The latter can only occur when the assignments to the X_i make all clauses

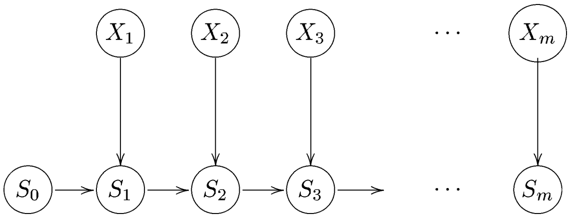

Fig. A.1. Extension to the network in Fig. 2.

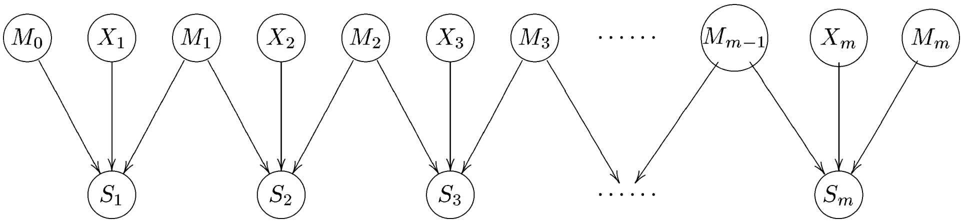

Fig. A.2. Bayes net with mediator nodes.

true, i.e., when there is at least one solution to the SAT problem. We need the factor δ in order to ensure that the probability of the evidence is strictly positive, even if the 3-SAT problem is unsatisfiable. Selecting $1/4^m \geqslant \delta > 0$ results in the posterior probability of S_m being greater than $\frac{1}{2}$ if there is a solution, and 0 if there is no solution, thus in this reduction we also get the "hard to approximate" result.

The network resulting from the above reduction, although (directed path) singly-connected network, is no longer a 2-level network. It is possible, instead of the ladder used in the previous reduction, to add additional binary-valued "satisfiability propagation" nodes S_i as well as additional binary-valued "mediator nodes" M_i (with $m \geqslant i \geqslant 0$) that collect information on the X_i in a 2-level network into node S_m (see Fig. A.2).[4] The mediator nodes M_i have uniform priors (i.e., 0.5 for each state). Each satisfiability propagation node S_i has as parents the nodes M_i, X_i, M_{i+1}. Node S_i has value T with conditional probability 1 just in the cases where information "propagates correctly" (see below)—and otherwise 0. Information is considered to "propagate correctly" just when $M_{i+1} = M_i = T$ and X_i is non-$\perp$, or when $M_{i+1} = F$ and either $X_i = \perp$ or $M_i = F$.

The state $S_i = T$, for all $m \geqslant i \geqslant 1$, is added as evidence (in addition to the $C_j = T$ evidence). Clearly, the additional evidence forces propagation of the state of the X_i variables—any non-zero posterior probability instantiation that has $M_m = T$ must also have all other $M_i = T$ and all X_i must be non-$\perp$. Thus, all instantiations with non-zero posterior probability that have $M_m = T$ are solutions to the 3-SAT problem, and $P(M_m = T) > 0$ iff the SAT instance is satisfiable.

References

[1] A. Abdelbar, S. Hedetniemi, Approximating MAPs for belief networks is NP-hard and other theorems, Artificial Intelligence 102 (1998) 21–38.

[2] R. Bubley, M. Dyer, Graph orientations with no sink and an approximation for a hard case of #SAT, in: Proceedings of Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA, 1997, pp. 248–257.

[3] E. Charniak, Bayesian networks without tears, AI Magazine 12 (4) (1991) 50–63.

[4] G.F. Cooper, The computational complexity of probabilistic inference using Bayesian belief networks, Artificial Intelligence 42 (2–3) (1990) 393–405.

[4] Independently, David Poole suggested in a personal communication that adding new evidence nodes can be used to "fold" the network in a similar manner—and presumably his scheme can be used to reduce any BN into a 2-level BN, resulting in yet another alternate proof.

[5] T.H. Cormen, C.E. Leiserson, R.L. Rivest, Introduction to Algorithms, MIT Press, Cambridge, MA, 1990.
[6] P. Dagum, M. Luby, Approximating probabilistic inference in Bayesian belief networks is NP-hard, Artificial Intelligence 60 (1) (1993) 141–153.
[7] A. Darwiche, Conditioning methods for exact and approximate inference in causal networks, in: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Montreal, QB, 1995, pp. 99–107.
[8] A. Darwiche, Recursive conditioning, Artificial Intelligence (Special Issue on Resource Bounded Reasoning) 125 (1–2) (2000) 5–41.
[9] S. Dasgupta, Learning polytrees, in: Proceedings of 15th International Conference on Uncertainty in Artificial Intelligence, Stockholm, Sweden, 1999, pp. 134–141.
[10] F.J. Diez, Local conditioning in Bayesian networks, Artificial Intelligence 87 (1–2) (1996) 1–20.
[11] C. Domshlak, R. Brafman, Structure and complexity in planning with unary operators, in: Proceedings of the Sixth International Conference on AI Planning and Scheduling, Toulouse, France, 2002, pp. 34–43.
[12] M.R. Garey, D.S. Johnson, Computers and Intractability, A Guide to the Theory of NP-completeness, W.H. Freeman, New York, 1979, p. 190.
[13] E.J. Horvitz, H.J. Suermondt, G.F. Cooper, Bounded conditioning: Flexible inference for decisions under scarce resources, in: Proc. 5th Workshop on Uncertainty in AI, Windsor, ON, 1989, pp. 182–193.
[14] F.V. Jensen, K.G. Olsen, S.K. Andersen, An algebra of Bayesian belief universes for knowledge-based systems, Networks 20 (1990) 637–660.
[15] S. Khuller, An $O(V^2)$ algorithm for single connectedness, Inform. Process. Lett. 72 (3–4) (1999) 105–107.
[16] J.H. Kim, J. Pearl, A computation model for causal and diagnostic reasoning in inference systems, in: Proceedings IJCAI-83, Karlsruhe, Germany, 1983, pp. 190–193.
[17] S.L. Lauritzen, D.J. Speigelhalter, Local computations with probabilities on graphical structures and their applications to expert systems, J. Royal Statist. Soc. 50 (1988) 157–224.
[18] V. Lepar, P.P. Shenoy, A comparison of Lauritzen–Spiegelhalter, Hugin, and Shenoy–Shafer architectures for computing marginals of probability distributions, in: Proceedings of 14th International Conference on Uncertainty in Artificial Intelligence, Madison, WI, 1998, pp. 328–337.
[19] R.E. Neapolitan, Probabilistic Reasoning in Expert Systems, Wiley, New York, 1990.
[20] J.D. Park, MAP complexity results and approximation methods, in: Proceedings of Eighteenth Conference on Uncertainty in AI, Edmonton, AB, Morgan Kaufmann, San Mateo, CA, 2002, pp. 388–396.
[21] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[22] Y. Peng, J.A. Reggia, A probabilistic causal model for diagnostic problem solving (parts 1 and 2), IEEE Trans. Systems Man Cybernet. SMC-17 (1987) 146–162, 395–406.
[23] D. Poole, Probabilistic Horn abduction and Bayesian networks, Artificial Intelligence 64 (1) (1993) 81–129.
[24] D. Poole, Probabilistic conflicts in a search algorithm for estimating posterior probabilities in Bayesian networks, Artificial Intelligence 88 (1–2) (1996) 69–100.
[25] D. Roth, On the hardness of approximate reasoning, Artificial Intelligence 82 (1–2) (1996) 273–302.
[26] E. Santos Jr., S.E. Shimony, Belief updating by enumerating high-probability independence-based assignments, in: Uncertainty in AI, Proceedings of the Tenth Conference, Seattle, WA, 1994, pp. 506–513.
[27] R.D. Shachter, Evaluating influence diagrams, Oper. Res. 34 (6) (1986) 871–882.
[28] S.E. Shimony, Finding MAPs for belief networks is NP-hard, Artificial Intelligence 68 (2) (1994) 399–410.
[29] E. Tsang, Foundations of Constraint Satisfaction, Academic Press, New York, 1993.
[30] L.G. Valiant, V.V. Vazirani, NP is as easy as detecting unique solutions, in: Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing, Providence, RI, 1985, pp. 458–463.
[31] Y. Weiss, Correctness of local probability propagation in graphical models with loops, Neural Comput. 12 (1) (2000) 1–41.